# Chapter 6: Data Management

*6.1 Overview of Data Management*

Collecting natural resource data is the first step toward understanding the ecosystems within the national parks. These ecosystems are evolving, as is the knowledge of them and how they work. Researchers use these "raw" data to analyze, synthesize, and model aspects of ecosystems. In turn, the results and interpretations are used to make decisions about the parks' vital natural resources. Thus, *data* collected by researchers and maintained through sound data management practices will become *information* through analyses, syntheses, and modeling. This transformation can only be achieved through the development of a modern information management infrastructure (e.g., staffing, hardware, software) and procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision making, research, and education.

This chapter summarizes the SFAN data management strategy, which is more fully presented in the SFAN Data Management Plan (DMP; Press 2005). The SFAN DMP serves as the overarching strategy for achieving the goals noted above. The plan supports I&M program goals and objectives by ensuring that program data are documented, secure, and remain accessible and useful indefinitely.

*6.1.1 Data Management Strategy*

The SFAN data management strategy holds that all data and derived information generated or otherwise used by the program will meet a high level of quality standards. Further, all data and information the SFAN program deems necessary to meet objectives, and that are not otherwise maintained, will be archived, documented, and made easily available and accessible. Data and information will be managed in a transparent manner such that all components may be easily compared by location, time and subject. Data and information will be accompanied by supporting documentation (metadata) that provide context, value, utility, and longevity, thereby facilitating broad understanding of SFAN program output to current and future end users.

The overarching goals of SFAN data management are to:

- ensure the highest quality and accuracy of program data
- fully qualify, document, and catalog all data to ensure their proper interpretation and use
- maintain data in an environment that ensures the long-term security and integrity of data
- ensure the longevity of data by keeping data formats standardized and current
- provide data in a variety of formats and venues to reach all potential users

The following objectives of the SFAN Data Management Plan help frame the strategy to meet SFAN data management goals:

Overall objectives:

- Outline the long-term goals of a comprehensive data management strategy for the SFAN I&M program
- Associate data management goals with the long-term goals of the network and service-wide I&M program
- Outline the procedures and work practices that support effective data management
- Guide current and future staff of the SFAN to ensure that sound data management practices are followed
- Guide the enhancement of legacy data to match formats and standards put forth in this plan
- Encourage effective data management practices as an integral part of project management so all data are available and usable for park management decisions now and into the future

Specific Objectives:

- Establish roles and responsibilities of SFAN program staff for managing data
- Identify necessary elements for a functional data management program and describe any anticipated changes to those elements
- Establish an organizational scheme for SFAN program data and information so that they are retrievable by staff, cooperators, and the public
- Establish basic quality control standards
- Establish standards for data, data distribution, and data archiving to ensure the long-term integrity of data, associated metadata, and any supporting information

*6.1.2 Types of Information Managed by the SFAN*

The term "data" is frequently used in a way that also encompasses other products that are generated alongside the tabular and spatial data that are the primary targets of data management efforts. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 6.1).

These data categories can contain one or more of the following data formats:

- hard-copy documents (e.g., reports, field notes, survey forms, maps, references, administrative documents)
- objects (e.g., specimens, samples, photographs, slides)
- electronic files (e.g., Word files, email, websites, digital images)
- electronic tabular data (e.g., databases, spreadsheets, tables, delimited files)
- spatial data (e.g., shapefiles, coverages, remote-sensing data)

Each of these data formats has specific requirements for ongoing management and maintenance, which are addressed in the SFAN DMP.

Table 6.1 Categories of data products and project deliverables.

| Category | Examples |
|---|---|
| Raw data | GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens |
| Compiled/derived data | Relational databases, tabular data files, GIS layers, maps, species checklists |
| Documentation | Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, NPS- Federal Geographic Data Committee (FGDC) standard metadata, database design documentation, quality assurance report, catalog of specimens/photographs |
| Reports | Annual progress report, final report (technical or general audience), periodic trend analysis report, publication |
| Administrative records | Contracts and agreements, study plan, research permit/application, other critical administrative correspondence |

*6.2 Data Management Roles and Responsibilities*

For the SFAN I&M program to work effectively, everyone within the network will have stewardship responsibilities in the production, analysis, management, and/or end use of data and information. The SFAN Data Management Plan specifies basic roles and responsibilities spanning the spectrum of data handling from collection to archiving. This spectrum includes field technicians, projects leaders, GIS specialists, and data managers. More detailed roles and responsibilities are given in the protocol for each monitoring project. Table 6.2 lists these basic roles and principal responsibilities.

Chief personnel involved with data management include project leaders and data managers. Figure 6.1 illustrates the core data management duties of the project leader and data manager and where those duties overlap. The Network coordinator interacts with project leaders to ensure that timelines for data entry, validation, verification, summarization/analysis and reporting are met.

*6.2.1 Project Leaders*

Project leaders oversee and supervise all phases of a monitoring project and are the point of contact for that project. Each project has two project leaders (one lead, one backup), which may consist of network, park or regional staff. They are responsible for the coordination and supervision of all phases of the project, from raw data collection to data validation and documentation to data analysis and reporting. They are also responsible for complying with the protocol methods and data management plan. They are responsible for the final submission of all products and deliverables. For projects involving contractors and/or cooperators, the project leader is also the Contracting Officers Technical Representative (COTR), and must insure that

Table 6.2 Summary or programmatic roles and responsibilities for data stewardship.

| Role | Data Stewardship Responsibilities |
|---|---|
| Network Coordinator | Ensure programmatic data and information management requirements are met as part of overall Network business.<br><br>Communicate with Network staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business. |
| Lead Data Manager | Serve as Point of Contact for National Park Service database applications (NPSpecies and NatureBib).<br><br>Communicate with national-level I&M Program for updates on NPS database applications and data standards. |
| Project Leader | Ensure useful data is collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting.<br><br>Develop, document and implement standard procedures for field data collection and data handling.<br><br>Supervise and certify all field operations.<br><br>Produce regular summary reports and conduct periodic trend analysis of data. |
| Project Data Manager | Develop and maintain the infrastructure for metadata creation, project documentation, and project data management.<br><br>Create and maintain project databases in accordance with best practices and current program standards.<br><br>Establish and implement procedures to protect sensitive data. |
| GIS Specialist | Coordinate and integrate local GIS and resource information management with Network, regional, and National standards and guidelines. |
| Project Technician | Record, enter and verify measurements and observations based on project objectives. |
| Information Technology/ Systems Specialist | Provide and maintain an information systems and technology foundation to support data management |
| I&M Data Manager (National Level) | Provide servicewide database support and services. |
| End Users (managers, scientists, public) | Provide necessary and requested feedback, review, and comments in order to sustain the continuous improvement of network operations and services. |

the contractor complies with the terms of the contract or cooperative agreement. Their active involvement in data management determines the quality and usefulness of the project data and overall success and longevity of the I&M Program.

*6.2.2 Data Manager*

Data Managers oversee the development, implementation, and maintenance of data infrastructure and standards for specific parks. Each I&M project is assigned a Data Manager for the duration of the project. Data Managers facilitate coordination between projects and protocols to allow for interchange of information wherever possible. Data Managers work with project leaders to design databases and software applications, facilitate data dissemination and coordinate long-term storage and maintenance of the data

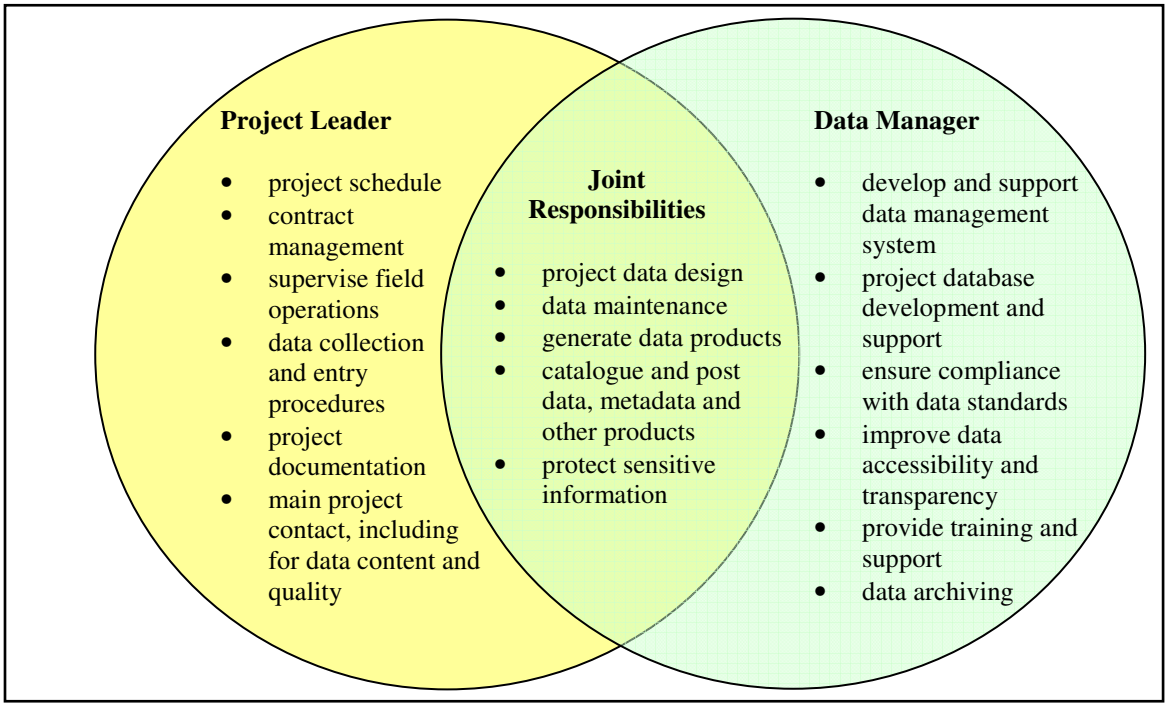


Figure 6.1 Core project data stewardship duties of project leaders and data managers

*6.3 Data Management Infrastructure*

The information system architecture necessary to fulfill the role of program data management includes national, regional, and park level infrastructure. Systems architecture signifies the applications, database systems, repositories, and software tools that make up the framework of the data management enterprise. The national level I&M data management infrastructure and strategy are used as a basis for data management in the SFAN.

*6.3.1 National Level I&M Data Management Infrastructure*

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and I&M staff integrate desktop database applications with internet-based databases to serve both local and national-level data and information requirements. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage.

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) used by the SFAN utilize a distributed application architecture with both desktop and internet-accessible (master) components (Figure 6.2). In addition, the SFAN has adopted relational database design standards in accordance with the Natural Resource Database Template, a relation database model developed in MS Access by the NRPC (Figure 6.2).

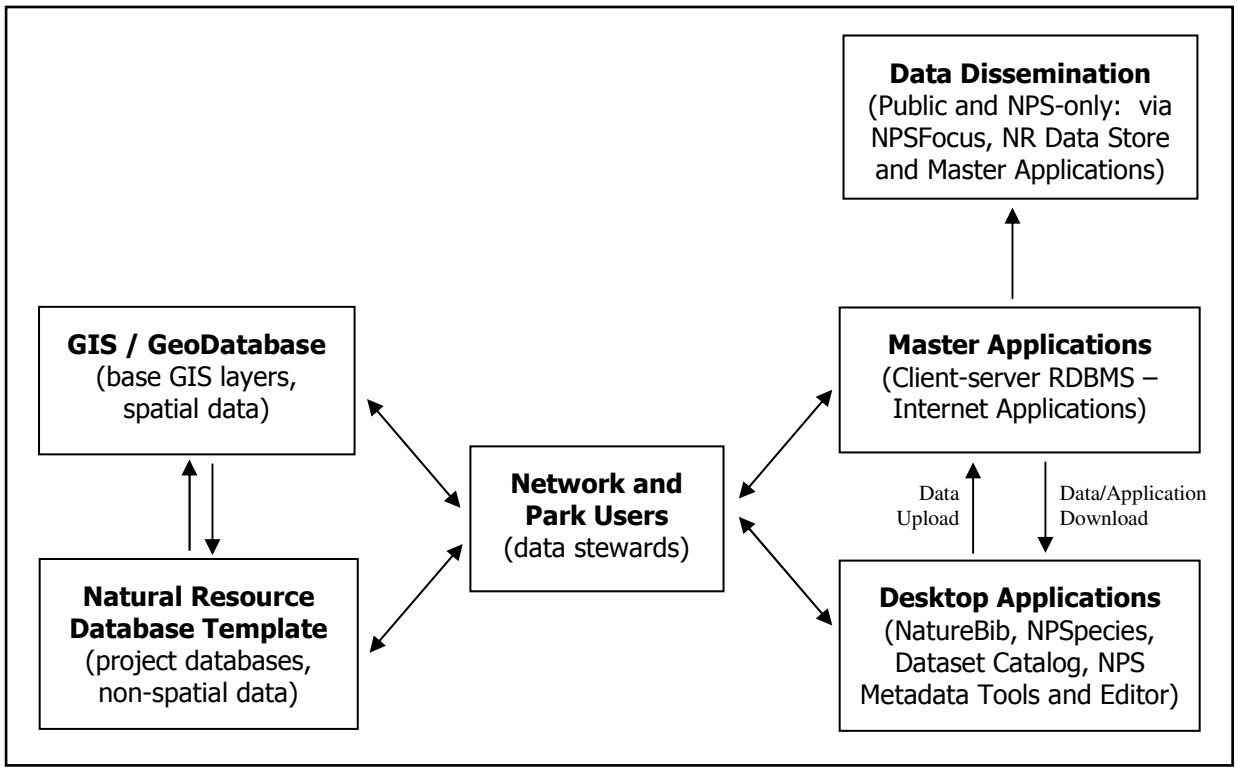


Figure 6.2 Model of the national-level application architecture

*6.3.2 Network and Park Level Data Management Infrastructure*

An important element of a data management program is a reliable, secure network of computers and servers. The SFAN digital infrastructure has three main components: servers maintained at the national level, park-based local area networks (LAN) nested within the Pacific West Region wide area network (WAN), and a Network directory nested within the GOGA LAN (Figure 6.3).

This infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups.

These components each host different parts of the natural resource information system.

**Park LANs**

- Local applications – desktop versions of national applications such as NPSpecies, Dataset Catalog, and NPS Metadata Tools and Editor
- Working files – working databases, draft geospatial themes, drafts of reports, administrative records
- Park digital archives – base spatial data, finalized datasets, and finished versions of park project deliverables
- Park GIS files – base spatial data, imagery, and project-specific themes

**Network Directory**

- Master project databases – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality
- Network digital archives – network repository for finished versions of project deliverables for I&M projects (e.g., reports, methods documentation, data files, metadata, etc.)

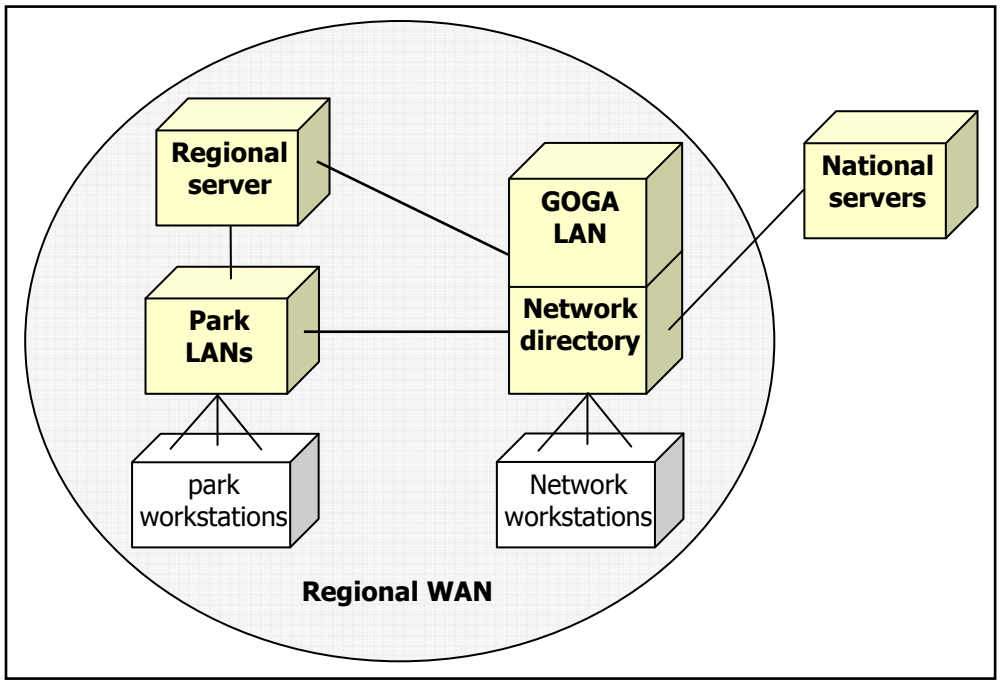


Figure 6.3 Schematic representing the layout and connectivity of SFAN computer resources

The Marin Headlands GOGA server is a critical component of the SFAN computer infrastructure. A separate directory on this server houses the Network directory, which includes all of the SFAN central files and digital archives. The server has power supply and storage

redundancy built to ensure data are kept safe. A tape drive is also connected to back up data on a regular weekly schedule. The last tape of the month is taken offsite to further protect the data.

*6.4 Data Management Standards*

The SFAN will conform to NPS standards and policy in all aspects of program data management operations in the interest of program integration and information sharing. The SFAN DMP specifies the standards by which data will be handled. Data management elements or principles common to more than one vital sign will be managed in a conventional manner to allow for greater comparison of data across the network, as well as to ensure further general data integrity.

*6.5 Data Acquisition and Processing*

The NPS I&M Program is responsible for acquiring the necessary information required by park managers to properly manage and maintain the natural resources of their park. To successfully accomplish this task, information from multiple sources is collected by the SFAN I&M Program and processed to ensure that it meets the data standards established by the SFAN. The DMP describes the general procedures the SFAN follows for acquiring and processing natural resource-related data. Procedures will vary depending on the data source, which can be placed into three general categories:

- **SFAN data:** data resulting from projects that are initiated, sponsored, or funded by the San Francisco Bay Area Network.

- **Other NPS data:** data resulting from projects that are initiated, sponsored, or funded by park units, or by regional or national NPS programs.

- **External data**: data produced or managed by agencies, organizations, or individuals other than the NPS.

The collection of programmatic data under the purview of the SFAN I&M Program is connected to either natural resources inventories or to vital signs monitoring. Data for each of these projects will enter and flow through the system illustrated in Figure 6.2.

Information and data sets available from other NPS (i.e., Exotic Plant Management Teams, Joint Fire Science Program) and external programs (USGS, NOAA) are utilized by the SFAN to strengthen and support its inventory and monitoring programs. These data sets can help to establish base resource conditions and aid in the detection of long-term monitoring trends.

*6.6 Quality Assurance/Quality Control*

The network will establish and document protocols for the identification and reduction of error at all stages in the data lifecycle. Although specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all network projects.

Each vital sign protocol will include specifics that address quality control. These may include:

- Field crew training
- Standardized data sheets
- Use of handheld computers
- Equipment maintenance and calibration
- Procedures for handling data (including specimens) in the field
- Data entry, verification and validation

Data entry after the field season represents a critical data life stage in terms of QA/QC. To facilitate data entry, data for each vital sign will be entered via customized MS Access applications modeled after the Natural Resources Database Template developed by the National I&M Program. These applications help enforce data standards by constraining the type, value, and format of data as appropriate to each vital sign.

The SFAN DMP presents several options for carrying out data verification (ensuring data on field sheets match data entered into a database) and validation (ensuring that the data make sense). Each vital sign protocol specifies procedures for completing proper verification and validation of data.

### *6.6.1 Documentation of Quality*

The final step in data QA/QC is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by each vital sign project leader and incorporated into formal metadata, as well as the SFAN primary data repository. Metadata for each data set/database will also provide information on the specific QA/QC procedures applied and the results of the review. Metadata and data will be available via the NPS NR-GIS Data Store.

## *6.7 Data Documentation*

Data documentation is a critical step towards ensuring that data are useable for its intended purposes well into the future. This involves the creation of metadata. Metadata can be defined as data about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog datasets, within intranet and internet systems, thus making these datasets available to a broad range of potential data users.

Metadata for all SFAN monitoring data will conform to the NPS Metadata Profile, which combines the FDGC standard, elements of the ESRI metadata profile, the Biological Data Profile, and NPS-specific elements (FDGC 1998). The SFAN metadata plan is limited to four recommended desktop applications for collecting metadata. These include Dataset Catalog and the NPS Metadata Tools and Editor, both developed by the NPS I&M Program, and two commercial off the shelf metadata tools, ArcCatalog and SMMS.

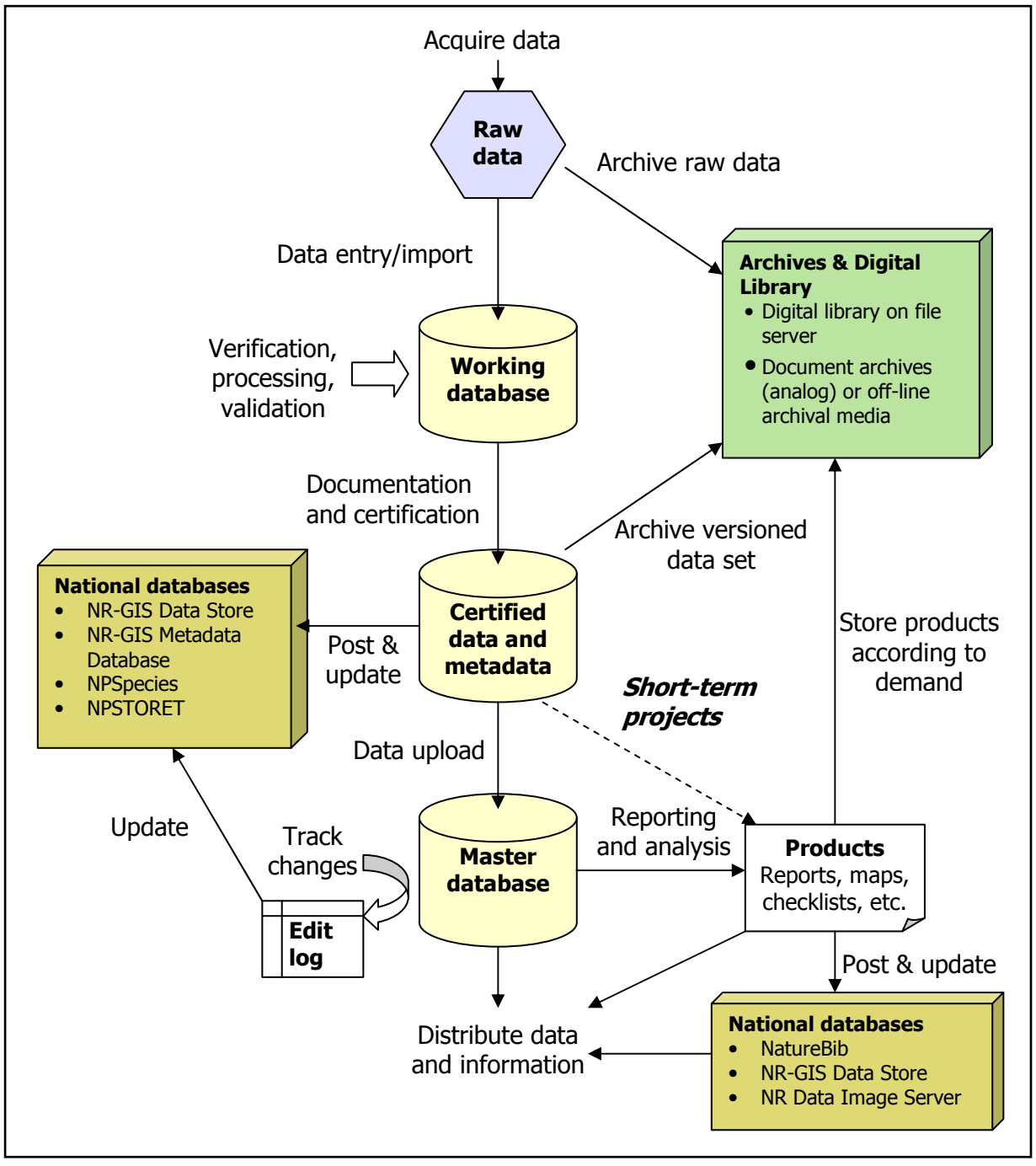


Figure 6.4 Diagram of the typical project data life cycle.

All relevant data products and associated metadata generated through the SFAN I&M Program will be posted to the NR-GIS Data Store. Launched in June 2005, the NR-GIS Data Store is a web-based system designed to integrate data dissemination and metadata maintenance for Natural Resource, GIS, and other program data sets, digital documents, and appropriate digital photos. The NR-GIS Data Store provides two functions: the NR-GIS Metadata Database and the NR-GIS Data Server. The NR-GIS Metadata Database is a repository of and search engine for metadata describing natural resource and GIS data. The NR-GIS Data Server hosts natural resource and GIS data (documented by the metadata in the NR-GIS Metadata Database) for download.

### *6.8 Data and Information Dissemination*

Access to SFAN monitoring products will be facilitated via a variety of data and information systems employing tools that allow potential users to browse, query, and obtain data, information, and supporting documents easily.

Providing well-documented data in a timely manner to park managers is especially important to the success of the program. The SFAN will make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control (legacy data, unknown data quality, Freedom of Information Act (FOIA) requests will be released with a disclaimer stating as such
- Distributed data are accompanied by metadata that clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate distribution
- A complete record of data distribution/dissemination is maintained

The network's main mechanism for distribution of I&M data will be the internet. Use of the internet will allow the dissemination of data and information to reach a broad community of users. As part of the NPS I&M Program, web-based applications and repositories have been developed to store a variety of park natural resource information. Table 6.3 outlines the applications and repositories that the SFAN will use to distribute data developed by the program.

#### *6.8.1 Data Ownership, FOIA, and Sensitive Data*

SFAN products are considered property of the NPS. However, the FOIA establishes a general right for any person to access federal agency records that are not protected from disclosure by any exemption or by special law enforcement record exclusions. The SFAN complies with all FOIA strictures regarding sensitive data. Each vital sign project leader, as the chief data steward, determines data sensitivity in light of federal law and stipulates conditions for release of the data in the project protocol and metadata.

Table 6.3 Repositories for SFAN products.

| Repository | Product |
|---|---|
| SFAN Digital Archive | Project data, metadata, and other products<br>Raw and certified data sets<br>Metadata, protocols, SOPs<br>Reports and administrative records<br>Digital photographs, derived products |
| SFAN Project Databases | Comprehensive data for multi-year projects |
| Park Collections, Museums, Herbariums, and/or National Archives | Administrative records, voucher specimens, raw data forms, hard copy reports |
| National Databases<br>- NPSTORET, NPSpecies, NatureBib | Compiled information about water quality, park species lists and taxonomic documentation, park resource bibliographies |
| NR Data Image Server | Copies of digital reports and other documents (catalogued in NatureBib) |
| NR-GIS Data Store | Metadata and data sets (spatial and non-spatial and products) |
| SFAN Website | Protocols, SOPs, and reports for all I&M data produced by the network. |

### *6.9 Data Maintenance, Storage, and Archiving*

The SFAN DMP describes procedures for the long-term management and maintenance of digital data, documents, and objects that result from SFAN projects and activities. The overall goals of these procedures are:

- to avert the loss of information over time
- to ensure that information is properly interpreted by a broad range of users
- to ensure that information can be easily obtained and shared through future decades

#### *6.9.1 Digital Data Maintenance*

In general, digital data maintained over the long term will be one of two types: short-term data sets, for which data collection and modification have been completed (i.e., inventory projects); and long-term monitoring data sets, for which data acquisition and entry will continue indefinitely.

Maintaining digital files involves managing the ever-changing associated infrastructure of hardware, software, file formats, and storage media. As software and hardware evolve, data sets must be consistently migrated to new platforms.

Data sets created or managed by the SFAN will be archived in read-only format with accompanying metadata, other data documentation, protocols, and final reports according to a

specified project schedule (monitoring) or upon project finalization (inventory). All finalized files will be stored in the SFAN archive directory on the GOGA Headlands server..

*6.9.2 Storage and Archiving Electronic Data and Documents*

Digital archives of completed I&M products, including SOPs, reports, and data sets, will be maintained at the park and Network levels. For long-term monitoring projects, data sets will be uploaded to both digital archive locations on an annual basis according to schedules outlined in the SOPs.

Final digital products, including SOPs, final reports, and data sets will be placed in a read-only format in the SFAN digital archives located on the Headlands server, Network I&M directory.

Only final documents will be archived – no drafts or works in progress. Only the lead data manager will archive data sets. Data sets must be validated and verified, must represent a complete set of records, and must have accompanying metadata and readme text files. Archived digital documents will then be entered online into NatureBib.

To ensure long term management of and protection for the work that is generated by the I&M program, a hardcopy of final products will be sent to the GOGA Records Center located in the Presidio of San Francisco and to individual park's archives as necessary.

*6.9.3 Storage and Archiving Hardcopy Documents and Objects*

Documents that are not available in digital format will either be scanned and saved as PDF files in the SFAN archive directory or saved in hardcopy format for larger documents. Scanned documents will then follow procedures outlined for electronic digital archives, which include entering the document into NatureBib and forwarding the original document to the GOGA Records Center and individual park's archives. Hardcopy documents will be maintained in a local library being developed at GOGA's Fort Cronkhite, with original copies forwarded to the GOGA Records Center and a record entered into NatureBib.

Specimens collected under the auspices of SFAN I&M program will be cataloged and maintained according to NPS Director's Order #24: NPS Museum Collections Management. Specific repositories for specimens are detailed in the inventory contracts or study plans, monitoring protocols and SOPs, and collection permits.

Several of the SFAN I&M projects incorporate digital, film (slides or negatives), and/or print photography into their protocols. Archiving procedures for digital and print photos will follow guidelines previously established for digital and hard-copy documents, respectively.

*6.10 Water Quality Data*

Water quality data collected to meet regulatory requirements is managed according to guidelines from the NPS WRD. This includes using the NPSTORET desktop database application at the parks to help manage data entry, documentation, and transfer. The network oversees the use of

NPSTORET according to the network's integrated water quality monitoring protocol and ensures the content is transferred at least annually to NPS WRD for upload to the STORET database maintained by the U.S. Environmental Protection Agency (Figure 6.5).

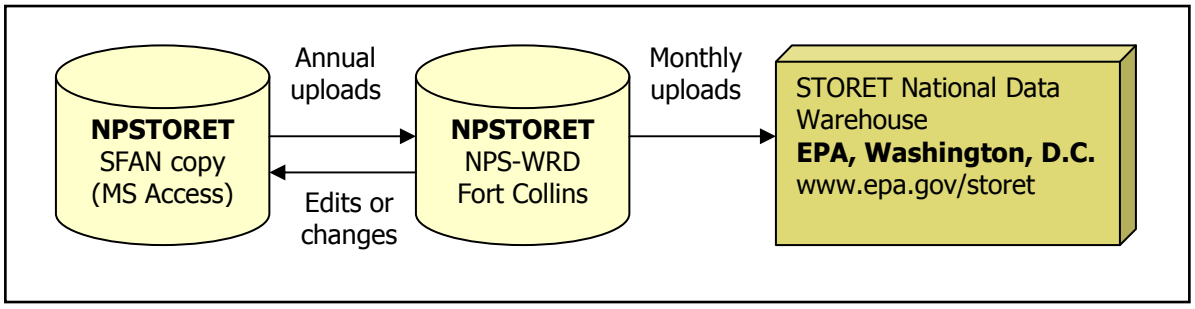


Figure 6.5 Data flow diagram for water quality data.